

# El análisis discursivo de fenómenos de violencia en un corpus de prensa digital

**CARMEN ÁVILA MARTÍN**

Profesora Titular  
Universidad de Granada  
Facultad de Filosofía y Letras  
Departamento de Lengua Española  
Campus de Cartuja s/n  
18071Granada  
E-mail: cavila@ugr.es

**RAMÓN MARTÍ SOLANO**

Universidad de Limoges  
Département d'Études Anglophones  
39E, Rue Camille Guérin  
87036 Limoges  
E-mail: ramon.marti-solano@unilim.fr

## EL ANÁLISIS DISCURSIVO DE FENÓMENOS DE VIOLENCIA EN UN CORPUS DE PRENSA DIGITAL

**RESUMEN:** El análisis del discurso mediático dispone en la actualidad de herramientas que proporcionan análisis empíricos cuantitativamente más documentados que los análisis tradicionales. El análisis léxico de las coocurrencias discursivas aporta datos de interés para interpretar cómo se construyen los discursos. Nuestro interés se sitúa en el empleo de la Lingüística de Corpus como instrumento para el Análisis del Discurso, pues ambas disciplinas presentan en la actualidad metodología adecuada para la interpretación de discursos como auxiliares en el análisis e interpretación de determinados fenómenos significativos. Para ello utilizaremos corpus específicos que nos aportan datos sobre la utilización discursiva de las unidades analizadas. Nos detendremos en determinadas manifestaciones de violencia en textos periodísticos que tienen denominaciones específicas en inglés y en español, no siempre coincidentes.

**PALABRAS CLAVES:** lingüística de corpus; análisis del discurso; lexicología; violencia; prensa.

**SUMARIO:** 1. Introducción. 2. El empleo de corpus en el análisis del discurso. 3. Metodología. 4. Fenómenos de violencia en la prensa inglesa. 5. Fenómenos de violencia en la prensa española. 6. Conclusiones.

## DISCOURSE ANALYSIS OF MANIFESTATIONS OF VIOLENCE IN A DIGITAL NEWSPAPER CORPUS

**ABSTRACT:** The analysis of media discourse has now tools that provide quantitative empirical analyses that are more documented than the traditional analyses. The lexical analysis of discursive co-occurrences provides interesting data that allow the interpretation of how discourses are constructed. We will focus on the use of corpus linguistics as an instrument for discourse analysis, as both disciplines currently present a methodology appropriate for the interpretation of discourses as auxiliary in the analysis and interpretation of certain significant phenomena. For this, we will use specific corpora that provide data on the discursive use of the units analysed. We will examine certain manifestations of violence in journalistic texts which have specific terms in English and in Spanish that do not always coincide.

**KEY WORDS:** corpus linguistics; discourse analysis; lexicology; violence; press.

**SUMMARY:** 1. Introduction. 2. The use of corpora in discourse analysis. 3. Methodology. 4. Manifestations of violence in the British press. 5. Manifestations of violence in the Spanish press. 6. Conclusions.

## ANALYSE DISCURSIVE DES PHÉNOMÈNES DE VIOLENCE DANS UN CORPUS DE PRESSE NUMÉRIQUE

**RÉSUMÉ:** L'analyse du discours a actuellement des outils disponibles qui permettent une analyse empirique quantitative beaucoup plus documentée que les analyses traditionnelles. L'analyse lexicale des co-occurrences discursives fournit des informations utiles pour interpréter la manière dont les discours sont construits. Notre intérêt est dans l'utilisation de la linguistique de corpus comme un outil pour l'analyse du discours, aussi bien que les deux disciplines présentent aujourd'hui une méthodologie appropriée pour l'analyse et l'interprétation de certains phénomènes significatifs. Nous utiliserons des corpus spécifiques qui fourniront des données sur l'usage discursif des unités testées. Nous nous intéresserons à certaines manifestations de violence dans la presse qui ont des dénominations spécifiques en anglais et en espagnol, pas toujours coïncidentes.

**MOTS CLÉS:** linguistique de corpus; analyse du discours; lexicologie; violence; presse.

**SOMMAIRE:** 1. Introduction. 2. L'utilisation de corpus en analyse du discours. 3. Méthodologie. 4. Phénomènes de violence dans la presse britannique. 5. Phénomènes de violence dans la presse espagnole. 6. Conclusions.

<b>Fecha de Recepción</b>	09/03/2013
<b>Fecha de Revisión</b>	04/11/2013
<b>Fecha de Aceptación</b>	05/11/2013
<b>Fecha de Publicación</b>	01/12/2013

## El análisis discursivo de fenómenos de violencia en un corpus de prensa digital

CARMEN ÁVILA MARTÍN & RAMÓN MARTÍ SOLANO

### 1. INTRODUCCIÓN

El presente trabajo se ocupa del análisis del tratamiento mediático de la violencia en la prensa británica y española a través del estudio de los elementos léxicos que la denominan. La alianza de la Lingüística de Corpus y el análisis textual permite obtener suficientes datos empíricos para explicar estos fenómenos y para situarlos en sus contextos interlingüísticos y extralingüísticos. Para ello realizaremos el estudio de algunas lexías recurrentes en la prensa británica y española. En el caso del inglés la lexía *hate crime* hace referencia a los delitos motivados por la hostilidad hacia la víctima como miembro de un grupo social. En la prensa española la utilización de expresiones de la violencia contra las minorías se manifiesta a través de la unidad léxica *acoso*. Para su estudio, se crea un corpus específico de cien mil palabras en cada lengua (inglés-español) y, después del análisis cuantitativo, se llevará a cabo un análisis cualitativo de los contextos léxicos más frecuentes de estas unidades léxicas. Al análisis cuantitativo de los datos lingüísticos procedentes de los corpus de prensa de los dos países le seguirá un análisis cualitativo y comparativo de los resultados obtenidos.

Este análisis se enmarca en el proyecto de investigación denominado “El tratamiento mediático de los fenómenos de violencia” de la Universidad de Limoges (Francia) dentro de la red internacional “Relations de genre et pratiques sociales: Amérique Latine, Europe, Caraïbes (ALEC)”. Se trata de un proyecto cuyo propósito es mostrar los contextos reales y las asociaciones lingüísticas relacionadas con actos violentos infligidos a mujeres, adolescentes, minusválidos, homosexuales y minorías religiosas o étnicas. Nos encontramos en una primera etapa de la investigación que será ampliada con un estudio comparativo del tratamiento mediático de diversas formas de violencia en la prensa española y de otros países. Partimos de la premisa de que la incidencia de ciertos tipos de violencia seguramente será más o menos importante en función de cada país y de sus características económicas, culturales o sociológicas. Sin duda, existirán formas de violencia más representadas y visibles, y por consiguiente más tratadas por los medios de comunicación, en ciertos países que en otros, pero solo un análisis riguroso permitirá establecer los límites de cada fenómeno concreto.

Resulta evidente que las sociedades modernas occidentales son en mayor o menor medida adictas a las imágenes de violencia: violencia en las pantallas de la televisión y del cine comercial, en los héroes de los cómics y en los personajes de los juegos de vídeo. En resumen, las distintas

manifestaciones de violencia forman parte de la cultura de masas contemporánea. Existe pues un bagaje común de representaciones culturales de los delitos y de la violencia pero también peculiaridades en cada sociedad que dan lugar a un cierto número de términos culturalmente marcados. Otro de nuestros objetivos consiste en señalar los estereotipos culturales y las representaciones culturales contemporáneas del delito y de la violencia y de sus diversas realizaciones léxicas.

## 2. EL EMPLEO DE CORPUS EN EL ANÁLISIS DEL DISCURSO

La justificación teórica y metodológica del enfoque tiene que ver en primer lugar con el desarrollo de la Lingüística de Corpus<sup>1</sup>. El acopio y procesamiento de datos lingüísticos permite la localización de unidades léxicas, el análisis de frecuencias, así como una fácil recuperación de los ejemplos que aportan información sobre los discursos. Nuestro interés por esta disciplina se debe a las posibilidades que ofrece para el análisis léxico. Partimos de la creación de corpus específicos según nuestro objeto de interés, y utilizamos estos corpus específicos (creados *ad hoc*) en comparación con los grandes corpus de las lenguas analizadas así como con los diccionarios generales.

Como segundo condicionamiento teórico, se tiene aquí en cuenta que las unidades léxicas son elementos privilegiados en la transmisión de los significados y contenidos ideológicos y culturales. El análisis de estas unidades léxicas no se realiza de forma aislada, sino en su aparición contextual y en un determinado género discursivo. Interesan también los significados, los contextos y las combinaciones más frecuentes, es decir, no solo las palabras aisladas, sino también las expresiones fijadas. De este modo el léxico adquiere en estos estudios un papel protagonista, el camino de acceso a los significados discursivos. En este sentido, la recuperación de las unidades léxicas como vía de acceso a los discursos constituye uno de los procedimientos del Análisis del Discurso, si bien no es el único. Este método de análisis debe ser completado con el análisis de otros mecanismos que confieren el sentido a los textos. Así, la intencionalidad comunicativa, las presuposiciones e inferencias o los significados retóricos (la ironía, la metáfora...) requieren que la comprensión de los discursos se lleve a cabo con análisis específicos en función de estas cuestiones.

Es innegable que este método ha despertado el interés de sociólogos, políticos o periodistas, que utilizan el análisis léxico para conocer los contenidos de textos con cierta repercusión social, como los debates políticos o los discursos institucionales. Desde una perspectiva lingüística

---

<sup>1</sup> Una primera versión del trabajo se presentó en el *III Congreso Internacional de Lingüística de Corpus* que tuvo lugar en Valencia en abril de 2011, con el título “El análisis discursivo de la violencia a través de un corpus específico”. En un trabajo similar se analizan las *lexias honour killing y violencia de género* (Ávila Martín y Martí Solano, 2012).

la Lingüística de Corpus, además de otras aplicaciones, tiene una larga trayectoria tanto en el análisis léxico como en el análisis discursivo. En ese sentido, se puede ver un estado de la cuestión y un recuento de trabajos en Baker (2006: 6 y ss.); una metodología similar se ha utilizado también en estudios de género (Del Teso, 2006). Por su parte, el discurso de la violencia ha sido objeto de atención desde la perspectiva de los estudios de cortesía (descortesía) en los medios de comunicación (C. Fuentes, 2010) con diferentes enfoques, aunque es difícil presentar “un paradigma claro y uniforme en su tratamiento” como señala esta autora en la presentación de los diferentes trabajos. El denominador común sin embargo es precisamente la preocupación por la presencia de la violencia verbal en los medios de comunicación y las manifestaciones de esa violencia en la sociedad actual a través de lenguaje.

Se hace preciso por tanto el análisis de la violencia verbal a través de herramientas que nos ayuden a objetivar la descripción de estos fenómenos y el empleo de las técnicas proporcionadas por la Lingüística de Corpus ayudan en su interpretación.

### 3. METODOLOGÍA

El método empleado se centra en las unidades léxicas, bien sea en sus colocaciones más frecuentes, o en el empleo de palabras clave, que también se pueden consultar en los grandes corpus generales, o en corpus especializados.

“However, there are also good reasons for building a specialized corpus: reference corpora may not contain enough of the text types you are interested in examining or may not have enough references to the subjects you want to investigate. And the data may be too old” (Baker: 2006, 31).

La creación de corpus especializados o específicos ofrece mayores garantías en la búsqueda las unidades léxicas que interesan ya que se trata de textos orientados con temáticas específicas. Las objeciones que se le hacen al estudio lingüístico basado en corpus tienen que ver, por un lado, con la representatividad del corpus, pero también con aspectos teóricos relacionados con la interpretación de los métodos cuantitativos. Desde el punto de vista teórico, sin embargo, la contraposición de métodos cuantitativos con métodos cualitativos se justifica en parte por el eclecticismo actual de los métodos en investigación. Las dificultades técnicas tampoco han sido ajenas a la creación de este tipo de estudios, pero, sin embargo, el desarrollo tecnológico posibilita en la actualidad el acceso a gran cantidad de datos con relativa facilidad, lo que permite la realización de análisis mucho más documentados que los realizados hasta este momento de desarrollo de los métodos cuantitativos.

Desde esta perspectiva, para este estudio se crean corpus específicos para el análisis de las lexías que interesan. En este sentido, cada una de

las lexías analizadas ha requerido la elaboración de su correspondiente corpus. El procedimiento se inicia con el rastreo en el motor de búsqueda de los archivos electrónicos del periódico de la lexía objeto de nuestro interés, con el fin de restringir los textos a artículos en los que aparece al menos una vez la lexía en cuestión. A continuación se crea una base de datos textual con estos artículos en los que aparece la unidad analizada hasta un número total de 100 000 palabras aproximadamente. Estos corpus, que a simple vista pueden parecer pequeños, son corpus específicos o temáticos, pero, sin embargo, en ellos las ocurrencias de las lexías que son objeto de análisis son muy numerosas si las comparamos con los resultados obtenidos en los corpus generales. El análisis de las lexías se centra en las concordancias, así como en las colocaciones y las listas de frecuencias que nos sirven para establecer las relaciones entre las unidades que nos interesan.

El programa utilizado para el procesamiento de los datos ha sido ConcGram 1.0. (C. Greaves, 2009) Esta herramienta permite identificar todas las coocurrencias de dos o más palabras independientemente de su forma gramatical y de su posición en el enunciado. La configuración de estas agrupaciones comienza con palabras contiguas a la derecha. Las líneas siguientes muestran ejemplos ordenados según la distancia entre las palabras de las unidades relacionadas. El programa indica también las frecuencias de las palabras, que junto con el análisis de palabras tema, las colocaciones y las palabras a las que están asociadas principalmente constituyen el objeto de nuestro estudio. Para este trabajo nos centramos en el análisis de dos lexías, que no presentan traducción exacta en las dos lenguas implicadas: la lexía *hate crime* en la prensa inglesa, que aunque puede aparecer traducida literalmente no es frecuente en español. Los fenómenos de violencia aparecen con frecuencia reflejados en la prensa española con la lexía *acoso*.

En este acercamiento a la metodología y análisis de corpus nos hemos interesado también por las relaciones entre las dos lenguas en la traducción y la denominación de los fenómenos de violencia como manifestaciones culturales. Si en algunos casos no hay una traducción exacta, consideramos, por otro lado, que se trata de fenómenos propios de cada cultura y que se expresan verbalmente de forma diferente. La dificultad de las metodologías que implican datos cuantitativos sigue estando en la interpretación que se haga de esos datos y en el uso que haga de ellos el investigador. Hemos considerado, sin embargo, que en una primera aproximación a estas cuestiones ya encontramos datos de interés que se ven corroborados por los métodos cuantitativos en su interpretación.

#### 4. FENÓMENOS DE VIOLENCIA EN LA PRENSA INGLESA

En la lectura de la prensa cotidiana británica se pueden detectar ciertas lexías, compuestas generalmente de dos sustantivos, que forman parte del campo semántico que podemos denominar “fenómenos violentos” o “actos de violencia”. Entre las cuales podemos citar las siguientes: *gang culture*, *hate crime*, *hoax bomb*, *honour killing*, *knife crime* y *white collar crime*. En primer lugar, comprobamos si estas lexías estaban registradas o no en diferentes diccionarios monolingües y bilingües, para lo cual utilizamos la tercera edición del *Cambridge Advanced Learner's Dictionary* en CD-ROM y llevamos a cabo una búsqueda automática de todos los sustantivos de la sección *Crime and Law*, lo que produjo un total de 946 entradas. Ninguna de estas lexías forma parte de la nomenclatura de este diccionario. A raíz de esta observación, se consultaron otros diccionarios para cotejar estos resultados: solo la lexía *hate crime* está lematizada en dos diccionarios electrónicos que son el *Oxford Advanced Learner's Dictionary* y el *Merriam-Webster's*, los dos diccionarios más recientes, lo que prueba la relativa novedad de esta lexía en inglés.

Nuestra segunda preocupación tenía que ver con la presencia de estas lexías en dos de los grandes corpus generales del inglés, el *British National Corpus* (BNC) y el *Corpus of Contemporary American English* (COCA). Los escasos, incluso a veces nulos, resultados demuestran por una parte que se trata de lexías de reciente creación y uso en el idioma inglés y por otra que, consecuentemente, solo los archivos electrónicos de un periódico de gran tirada pueden rendir cuenta de fenómenos léxicos de esta índole.

La relación entre el empleo de estas lexías y las noticias periodísticas sobre sucesos es obvia. Por esta razón decidimos crear nuestro corpus a partir de fuentes periodísticas. Es evidente que las lexías referidas a fenómenos de violencia se tratan también en otros géneros textuales especializados, como son los textos legales y los textos criminológicos, que pueden ser objeto de una investigación posterior.

Según los resultados obtenidos en los archivos electrónicos conjuntos del diario británico *The Guardian*, las lexías *hate crime*, *gang culture*, *honour killing* y *knife crime* son las más representadas, con *knife crime* a la cabeza y con un total de mil ocurrencias. El uso de *knife crime* es significativamente más frecuente en plural que en singular, a diferencia de lo que ocurre con *hate crime* en la que la diferencia en el número gramatical no es pertinente. Estas lexías no están prácticamente representadas en el *British National Corpus*, lo que convierte a este corpus general en una herramienta inapropiada para este tipo de estudio. Es muy significativo que tres de las cuatro lexías más frecuentes en los archivos periodísticos (es decir, *hate crime*, *gang culture* y *hounor killing*), no tienen ni una sola ocurrencia en el BNC. Si tenemos en cuenta que la base de datos textual del *British National Corpus* llega solo hasta el año 1993, podemos concluir, en vista de los resultados, que no solo se trata de lexías

que se han acuñado recientemente, sino que también se trata aquí de un rebrote de nuevas formas de violencia y de un uso a veces desmedido de estas cuestiones por parte de la prensa, sobre todo en el último decenio. Con respecto al COCA, y si comparamos los numerosos resultados de *hate crime* y *white-collar crime* y los prácticamente inexistentes (no hay que olvidar aquí que se trata de un corpus de 410 millones de palabras) de *honour killing* y *knife crime*, las conclusiones parecen evidentes: los primeros serían originalmente fenómenos culturalmente específicos de los Estados Unidos de Norteamérica y los segundos, del Reino Unido. El gran número de ocurrencias de *hate crime* en el COCA es un signo evidente de que se trata de una acuñación propia de la variedad del inglés americano, que se ha exportado fácilmente al otro lado del Atlántico, como lo demuestran los resultados en los archivos periodísticos, que la sitúan en segundo lugar después de *knife crime*, este último, un fenómeno, típicamente británico. Si exceptuamos los artículos, algunas preposiciones como *of* (de), algunas conjunciones como *and* (y) y las diferentes formas de los verbos *be* y *have* (ser o estar y haber o tener), que son las palabras más frecuentes en nuestros corpus, hemos seleccionado los sustantivos y adjetivos más frecuentes y, por ende, más significativos contextualmente.

Compounds	The British National Corpus (BNC)	Corpus of Contemporary American English COCA	<i>The Guardian</i>
<b><i>hate crime(s)</i></b>	0 + 0	<b>417 + 568</b>	<b>419 + 340</b>
<b><i>hoax bomb(s)</i></b>	<b>8 + 1</b>	<b>1 + 0</b>	<b>54 + 26</b>
<b><i>gang culture(s)</i></b>	0 + 0	13 + 0	<b>428 + 5</b>
<i>white(-)collar crime(s)</i>	<b>17 + 3</b>	<b>164 + 31</b>	159 + 24
<b><i>honour killing(s)</i></b>	0 + 0	0 + 2	<b>203 + 293</b>
<b><i>knife crime(s)</i></b>	4 + 1	3 + 1	<b>959 + 41</b>

Tabla 2.1. Resultados en los corpus ingleses y en el periódico *The Guardian*

#### 4.1. HATE CRIME

No resulta fácil definir *hate crime*. Utilizaremos aquí la definición propuesta en el informe del grupo de trabajo sobre *hate crime* del parlamento escocés, que lo define como “un delito motivado por un sentimiento de maldad o de hostilidad hacia un grupo social”. Según la versión electrónica del diccionario Merriam-Webster’s, el primer uso atestiguado de esta lexía se remontaría a 1984.

Se encontraron 410 artículos en los que aparece *hate crime* en los archivos del periódico para el primer decenio de este siglo. El mayor número de ocurrencias de esta lexía corresponde al año 2009. Cuenta con

186 ocurrencias en artículos procedentes en su inmensa mayoría de las secciones “UK news” (nacional) y “World news” (internacional).

Los primeros resultados corresponden a la lista de palabras más frecuentes:

people	404
police	296
crime	278
hate	220
gay	216
community	113
death	88
BNP	82
racist	75

Tabla 2.2. Palabras más frecuentes en el corpus de *hate crime*

Si exceptuamos por un lado las palabras *people* (personas) y *police* (policía) que aparecen de forma sistemática y con un elevado número de ocurrencias en todos nuestros corpus y que por esta razón no añaden elementos informativos específicos, y por otro lado los dos elementos constitutivos de la lexía en cuestión, o sea *hate* (odio) y *crime* (delito), la palabras más frecuentes por este orden son las siguientes: *gay* (homosexual), con 216 ocurrencias, *community* (comunidad), con 113 ocurrencias, *death* (muerte), con 88 ocurrencias y *BNP*, las siglas del *British National Party* (Partido Nacional Británico, formación política de extrema derecha), con 82 ocurrencias.

Pensamos que la relación que podría establecerse entre el aumento de este tipo de delito y la fulgurante ascensión del partido británico de extrema derecha, solo podría realizarse con un corpus mayor y con un análisis de coocurrencias aún más específico en el que se examinarían las posibles combinaciones e interrelaciones entre *hate crime* y *BNP*.

1	Sam Tarry, a campaign organiser for Hope Not Hate, said: "The BNP getting in has legitimised racist
2	more they are driven into our arms." Rise in hate crime follows BNP council election victories *
3	in the past six years. In eight wards reports of hate crime rose following BNP election wins despite a
4	between the rise of racist and Islamophobic hate crime and the ascendancy of the BNP." It is a
5	would review its policies on prosecuting race hate crimes following the election of two BNP
6	legitimised." The BNP denies that increases in hate crime are related to its activities and blames
7	seat for the BNP in May 2006. A year later, hate crime rose again and 45 racial incidents were
8	the BNP, racism and the police, and reducing hate crime. "Overt racism is not a problem on campus,
9	the BNP in May 2006, while in Tilbury, Essex, hate crime rose almost five-fold in the five years
10	a BNP member, in spite of declines in reported hate crime in the wider police areas. In other wards
11	BNP leaflets • Pressure mounts for end to race hate law loopholes • CPS powerless to pursue

Tabla 2.3. Concordancias de *hate/BNP*

Una segunda investigación se centra en las colocaciones léxicas. Nuestro objetivo es precisar las unidades léxicas situadas a la izquierda de esta base para así delimitar y cuantificar los diferentes tipos de *hate crime*. Los resultados muestran una evolución en el tratamiento de este fenómeno. La ley sobre el odio religioso y racial de 2006 del parlamento



británico definía *hate crime* en relación con la hostilidad hacia diferentes grupos étnicos y religiosos. Nuestra investigación demuestra que el tratamiento mediático de este fenómeno lo pone en relación principalmente con homosexuales y minusválidos en primer lugar, y que la relación con las minorías raciales y religiosas aparece claramente en segundo lugar.

homophobic/anti-gay	11
disability	11
racist/race	4
islamophobic/anti-Muslim	3
antisemitic/anti-Jewish	3
religious	1

Tabla 2.4. Colocaciones en el corpus de *hate crime*

En definitiva, estas lexías son de reciente creación y presentan una mayor frecuencia de uso en los archivos electrónicos del periódico en los años 2008 y 2009. Consideramos que existen esquemas que conforman y delimitan la representación cultural de los diferentes fenómenos de violencia y que estos esquemas se construyen a base de lexías, combinaciones, colocaciones y asociaciones léxicas.

## 5. FENÓMENOS DE VIOLENCIA EN LA PRENSA ESPAÑOLA

El estudio de los fenómenos de violencia en español nos lleva en primer lugar a determinar las denominaciones más frecuentes. La propia definición de la palabra *violencia* en el DRAE nos da un primer dato revelador, pues entre las definiciones mantiene “Acción de violar a una mujer”, donde se constata que la violencia es una acción que se ha ejercido históricamente contra las mujeres, aunque es una acepción que no se usa en la actualidad.

La violencia contra la mujer tiene su expresión en la lexía *violencia de género* que se utiliza con más frecuencia que la lexía *violencia doméstica* propiciada por la RAE. Su tratamiento en la prensa española es frecuente, aunque es un neologismo que no se incluye aún en los grandes diccionarios del español. No aparece ni en *Clave* (1996) ni en el diccionario REDES (2005). En la consulta del CREA (Corpus de Referencia del Español Actual), se encuentran 60 ocurrencias en un total de 32 documentos. Este corpus está compuesto, según los datos de la RAE, por más de 160 millones de palabras (hasta mayo de 2008). En la estadística que nos proporciona este corpus, la primera aparición es de 1993, en dos casos, y el número de casos aumenta hasta la presencia de 22 casos en 2004.

En cambio, en la creación de un corpus específico (del periódico *El País* entre 2010 y 2011) hemos recogido para esta lexía 341 casos para *violencia de género*, y aparecen otras lexías tales como *violencia machista* (78), *violencia doméstica* (10), *violencia sobre la mujer* (9), *violencia contra la mujer* (6) y *violencia sexista* (6), usadas todas como sinónimos de *violencia de género* pero con una frecuencia menor. La menor aparición de *violencia*

doméstica indica que la denominación *violencia de género*, a pesar de las críticas que se hicieron en su momento, alcanza una mayor difusión en esos años y en un tipo de prensa, dato que habría que comparar con otras fuentes (vid. Moreno 2010).

Estadísticas								
Año	%	Casos	País	%	Casos	Tema	%	Casos
2004	36.66	22	ESPAÑA	88.33	53	2.- Ciencias sociales, creencias y pensamiento.	53.33	32
2003	25.00	15	MÉXICO	3.33	2	3.- Política, economía, comercio y finanzas.	26.66	16
2001	13.33	8	NICARAGUA	3.33	2	4.- Artes.	8.33	5
2002	11.66	7	CHILE	1.66	1	6.- Salud.	8.33	5
2000	10.00	6	COSTA RICA	1.66	1	5.- Ocio, vida cotidiana.	3.33	2
1993	3.33	2	ECUADOR	1.66	1			

Tabla estadística 3.1. Aparición de *violencia de género* en el CREA

La violencia con respecto a otras minorías étnicas o sexuales se ha denominado *discurso del odio* o *crimen de odio*. Ambas expresiones, que proceden del inglés, no tienen una gran presencia en el discurso mediático, aunque se pueden encontrar algunos ejemplos. En el caso de *discurso de odio* hemos podido crear un corpus reducido de 16 000 palabras del periódico *El País*. En este caso, la aparición del *discurso del odio*, es un neologismo que todavía no ha adquirido carta de naturaleza; solo aparece en cinco ocasiones y en una ocasión está entre comillas.

1	su editorial a alertar sobre, así lo nombró, el <b>discurso de odio</b> . El lunes coronó la cadena de entuertos
2	Presentar como víctimas a los promotores de este <b>discurso de odio</b> dice mucho del estado de confusión
3	entre la realidad virtual creada por el <b>discurso del odio</b> y el despliegue efectivo de la violencia
4	acusa de propagar la doctrina nazi y el llamado " <b>discurso del odio</b> ". El escrito del fiscal dedica 40
5	no funcionan a golpe de intuiciones, que el <b>discurso del odio</b> no se imponga, y que las decisiones
6	por incitación al <b>odio</b> religioso, pero moderó su <b>discurso</b> político y fundó el AKP en 2001 para arrollar en

Tabla 3.2. Concordancias de *discurso de odio* en *El País*

En el caso de *crímenes de odio* los ejemplos son aún más reducidos. Sólo hemos podido recoger un corpus de unas 1000 palabras en el que esta expresión solo aparece en dos ocasiones.

1	frentes. Lideró la denuncia de los <b>crímenes de odio</b> por homofobia en una comisión ciudadana que hizo
2	gays y transexuales en <b>crímenes</b> motivados por el <b>odio</b> , equiparando homofobia y racismo. En enero, en el

Tabla 3.3. Concordancias de *crímenes de odio* en *El País*

En ambos casos se trata de préstamos no asentados en español, aunque se puedan encontrar algunos ejemplos de uso. Muchos de los fenómenos violentos con los que se relacionan estas unidades (homofobia y racismo), reciben la denominación de violencia seguida de un adjetivo o sintagma que la completa (por ejemplo, violencia religiosa, violencia política, -terrorismo-, violencia de género, violencia contra los homosexuales, etc.). Una forma de violencia tipificada como delito es el *acoso*, que va a constituir el objetivo de nuestro análisis.

### 5.1. Acoso

Partimos en este caso de una lexía simple que está aumentando las especificaciones en su significado. Es un término general, cuya única variante recogida en el DRAE (2001) es la de *acoso sexual* que se define como “El que tiene por objeto obtener los favores sexuales de una persona cuando quien lo realiza se halla en posición de superioridad respecto de quien lo sufre” (DRAE).

Por su parte, el diccionario REDES (2005) incluye toda una serie de adjetivos, algunos de los cuales nos remiten a formas de violencia específica:

**acoso.** asfixiante, constante, despiadado, frenético, fuerte, implacable, insufrible, obsesivo, periodístico, pertinaz, policial, salvaje, sexual, sin tregua, sofocante, tenaz, violento (...).

Los acosos *sexual, laboral, psicológico e inmobiliario* están tipificados como delitos en el Código Penal. El análisis de las expresiones de la violencia en el ámbito jurídico es otra de las vías de análisis de los discursos de la violencia. Sin embargo, en la etapa actual de nuestro trabajo nos centraremos en el análisis de un corpus específico periodístico creado para tal fin.

Después de un rastreo inicial con la lexía *acoso*, nuestro análisis se basa en la elaboración de un corpus de 100 000 palabras del periódico *El País* con un total de 155 artículos seleccionados sobre un total de 1153 publicados en el último año (del 25 de marzo de 2010 al 25 de marzo de 2011). En estos artículos, una vez analizados mediante el programa “Congram 0.1”, obtenemos los siguientes resultados: la unidad léxica *acoso* aparece en 242 casos. El compuesto léxico más utilizado es *acoso sexual* (37 ocurrencias), seguido de *acoso escolar* (14), *acoso laboral* (9), y *acoso moral* (4). No ha aparecido ningún caso de *acoso inmobiliario* en el último año, el año de la crisis inmobiliaria, pues es de suponer que ha cesado de producirse este acoso. Tampoco ha aparecido ningún caso de *acoso psicológico*.

Las unidades léxicas que presentan un mayor índice de frecuencia son las siguientes: *acoso* 242 ocurrencias, *contra* 229, *años* 218, *gobierno* 152, *Estado* 106, *presidente* 88, *política* 88, *mujeres* 83, *policía* 65 y *menores* 53.

La preposición *contra* remite al objeto del acoso. En las concordancias podemos obtener datos de las colocaciones más frecuentes (*medidas, asociación, consignas...*) *contra* el acoso y *contra* quien se ejerce el acoso (*contra* alguien o algo), así como fraseología, *acoso y derribo contra, campaña de acoso*.

1	deuda de los países en problemas. A pesar del <i>acoso</i> , el <i>Gobierno</i> de Portugal, que también se ve
2	el último y más grave episodio de una campaña de <i>acoso</i> <i>contra</i> el <i>Gobierno</i> de Murcia que se ha venido
3	que en ocasiones parecían más bien un plan de <i>acoso</i> y derribo al <i>Gobierno</i> de la Autoridad Palestina.
4	dice el PP, bien; si se hace lo que no le gusta, <i>acoso</i> y derribo <i>contra</i> el <i>Gobierno</i> . En materia
5	dice el PP, bien; si se hace lo que no le gusta, <i>acoso</i> y derribo <i>contra</i> el <i>Gobierno</i> . ¿No habíamos
6	agresión", que enmarcó en la "campaña de <i>acoso</i> " que cree que está sufriendo el <i>Gobierno</i> murciano
7	2002 tras el escándalo desatado por el caso de <i>acoso</i> sexual a la ex edil de su equipo de <i>Gobierno</i> , fue
8	en la Antártida El <i>Gobierno</i> nipón cede ante el <i>acoso</i> de la organización ecologista Sea Shepherd <i>contra</i>

Tabla 3.3. Concordancias *acoso/Gobierno*

El sustantivo más frecuente en el discurso después de la preposición es *años*, que aparece en 218 ocasiones. La búsqueda de los contextos en los que aparece esta unidad nos remite a la elaboración de las narraciones periodísticas que tienen en cuenta los elementos temporales de la noticia, la edad de los afectados, la duración de la condena impuesta o la cuantificación temporal del acoso, que puede durar años.

Entre las lexías más frecuentes son abundantes las relacionadas con las instituciones políticas: *gobierno* en 152 casos, *presidente* (88) y *política* (88), *estado* (84 del total de 107 ocurrencias, pues en 23 ocasiones no funciona como sustantivo sino como participio verbal) y *policía* con 65 apariciones. La coocurrencia de estas lexías pertenecientes al lenguaje institucional con *acoso* relacionan estas violencias con el discurso político presente en la prensa. Los casos de acoso están relacionados con esferas de poder que implican un abuso de las posiciones de fuerza que pueden ejercerse en la sociedad. Al mismo tiempo la prensa reproduce el discurso político y se constituye en portavoz de ese discurso que se difunde a través de los medios de comunicación.

Después de las lexías institucionales, la más frecuente es la que vincula el acoso a la mujer, aunque no es siempre la protagonista del acoso sexual, pues también se vincula a los menores.

1	países del mundo y que pretende denunciar el <b>acoso</b> a <b>mujeres</b> , un concepto que incluye los
2	de la denuncia por coacción, violación y <b>acoso</b> sexual de dos <b>mujeres</b> con las que mantuvo
3	fue declarado ayer culpable de doble violación y <b>acoso</b> sexual a tres <b>mujeres</b> que trabajaron con él
4	diciembre de doble violación a una empleada y de <b>acoso</b> sexual a otras tres <b>mujeres</b> que trabajaron con él
5	año recibe unas 3. 200 demandas por casos de <b>acoso</b> sexual y violación, en su mayoría a <b>mujeres</b> , pero
6	En Egipto, el 83% de las <b>mujeres</b> han sufrido <b>acoso</b> sexual alguna vez. Un triángulo diabólico formado
7	

Tabla 3.4. Concordancias acoso/mujeres

1	secretaría. 107. Alerta por el aumento de <b>acoso</b> a <b>menores</b> en la Red Policía y Guardia Civil
2	EL PAÍS   EFE - Madrid - 07 02 2011 Vota El <b>acoso</b> sexual a <b>menores</b> a través de internet está
3	graves daños psicológicos a los <b>menores</b> . El <b>acoso</b> sexual o grooming afecta a ambos sexos, a <b>menores</b>
4	de seguridad llaman a los <b>menores</b> a denunciar el <b>acoso</b> sexual en internet La mayoría de las víctimas de
5	ha arrestado a dos <b>menores</b> por un supuesto <b>acoso</b> moral a una compañera de clase, una alumna
6	que los <b>menores</b> denuncien si son víctimas de <b>acoso</b> sexual en internet. España es uno de los países
7	de los <b>menores</b> que indiquen que están sufriendo <b>acoso</b> 119. ENTREVISTA: SHIRIN EBADI Premio Nobel de la

Tabla 4.5. Concordancias acoso/menores

6. CONCLUSION

La recopilación de datos utilizando las herramientas que proporciona la Lingüística de Corpus permite la realización de análisis empíricos cuantitativamente mejor documentados que los análisis tradicionales. Los análisis presentados no agotan las posibilidades de interpretación que proporcionan estas herramientas, pero constituyen una muestra. Las unidades léxicas son la vía de entrada al discurso, a partir de las cuales se pueden realizar lecturas sobre la intencionalidad comunicativa y las características discursivas, pues el análisis léxico aporta datos de interés para interpretar cómo se construyen los discursos, aunque se deberán tener en cuenta otros aspectos.

La creación de corpus específicos se muestra como una metodología adecuada para ampliar los datos presentados en los diccionarios y los grandes corpus, y permite focalizar de forma específica las características de un determinado tipo textual. La creación de un corpus específico para cada unidad léxica analizada permite la obtención de datos que podrán ser interpretados cualitativamente en cada caso. Las lexías que se han seleccionado presentan un interés contrastivo entre las dos lenguas y en algunos casos plantean dificultades de traducción.

Finalmente, la expresión de los fenómenos de violencia es un hecho generalizado en las sociedades modernas, pero presenta peculiaridades en

cuanto a sus manifestaciones culturales y lingüísticas, por lo que otro de nuestros objetivos es poner de manifiesto esas diferencias culturales. En inglés *hate crime* se relaciona principalmente con homosexualidad y en segundo lugar con racismo, mientras que en español todavía no ha calado la denominación de *delitos de odio*. La aparición en la prensa de la lexía *acoso* está vinculada a grupos de poder y a la violencia sexual centrada exclusivamente en mujeres y niños.

## REFERENCIAS

- ÁVILA MARTÍN, C. & MARTÍ SOLANO, R. (2012): "Les phénomènes de violence dans la presse écrite au Royaume Uni et en Espagne", *DIRE*, 3, en <<http://epublications.unilim.fr/revues/dire/319>>.
- BAKER (2006): *Using corpora in Discourse Analysis*, Continuum: London- New York.
- BOSQUE, I. (2004): *Redes: Diccionario combinatorio del español contemporáneo: las palabras en su contexto*, Ediciones SM: Madrid.
- CHAKRABORTI, N. & GARLAND, J. (2009): *Hate Crime: Impact, Causes and Responses*, London: Sage.
- DALBERA J.P. (2002): "Le corpus entre données, analyse et théorie", *Corpus*, 1, 2002. <<http://corpus.revues.org/10>> [15/12/2003].
- FILLMORE, C. J. & ATKINS, B. T. S. (1994): "Starting where the Dictionaries Stop: The Challenge of Corpus Lexicography" in B. T. Atkins and A. Zampolli (eds.) *Computational Approaches to the Lexicon*, Oxford: Oxford University Press, pp. 349-393.
- FUENTES, C. & ALCAIDE LARA, E. (2008): *Des (cortesía), agresividad y violencia verbal en la sociedad actual*, Universidad Internacional de Andalucía.
- FUENTES, C. (coord.) (2010): "Violencia y discurso", *Discurso & Sociedad*, Vol. 4 (4).
- GREAVES, C. (2009): *ConcGram 1.0. A phraseological search engine*, Amsterdam: John Benjamins, Publishing Company. (CD-ROM).
- HALLIDAY, M. A. K. (2004): *Lexicology and Corpus Linguistics*. London: Continuum.
- HEROLD, A. (2007): "Corpus queries" in C. Fellbaum (ed.) *Idioms and Collocations. Corpus-based Linguistic and Lexicographic Studies*. London: Continuum, pp. 54-63.
- MINUGH, D. (1999): "You people use such weird expressions: the frequency of idioms in newspaper CDs as corpora", *Corpora Galore: Analyses and Techniques in Describing English*. Amsterdam: Rodopi, pp. 57-71.
- MORENO BENÍTEZ, D. (2010): "De violencia doméstica a terrorismo machista: el uso argumentativo de las denominaciones en la prensa", *Discurso & Sociedad*, vol. 4 (4), pp. 893-917.
- PARTINGTON, A. (1993): "Corpus Evidence of Language Change", *Text and Technology. In Honour of John Sinclair*, Amsterdam/Philadelphia: John Benjamins, pp. 177-192.
- REAL ACADEMIA ESPAÑOLA (2001): *Diccionario de la lengua española*, Real Academia Española: Madrid.
- REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA) [en línea]. *Corpus de referencia del español*

- actual*. <<http://www.rae.es>> [8/02/2011].
- SECO, M., ANDRÉS, O. y RAMOS, G. (1999): *Diccionario del español actual*, Aguilar, Madrid.
- SINCLAIR, J. (2004): *Trust the text- Language, Corpus and Discourse*. London: Routledge.
- STUBBS, M. (1996): *Text and Corpus Analysis Computer assisted Studies of Language and Culture*. Oxford: Blackwell.
- STUBBS, M. (2001): *Word and phrases. Corpus studies of Lexical Semantic*. Oxford: Blackwell.
- TESO CRAVIOTTO, M. (2006): "Words that matter: Lexical choice and gender ideologies in women's magazines", *Journal of Pragmatics*, 38, pp. 2003-2021.
- VV. AA. (2008): *Cambridge Advanced Learner's Dictionary*, Third Edition (CALD3) on CD-ROM.
- VV. AA. (1996): *Clave: diccionario de uso del español actual*, SM, Madrid.
- VV. AA. (2004): *Oxford Advanced Learner's Dictionary*, (6ª ed.) <[www.oxfordadvancedlearnersdictionary.com](http://www.oxfordadvancedlearnersdictionary.com)>.
- VV. AA. (2007): *The Oxford-Hachette French Dictionary*, (4ª ed.), [www.merriam-webster.com/dictionary](http://www.merriam-webster.com/dictionary).
- VV. AA. (2004): *Working Group on Hate Crime Report*, Scottish Executive.